

Informe de Adaptación de DSpace a Europeana

Fase Danubio Europeana Data Model (EDM)

Parte II: Propuesta de solución técnica

Elaborado por: *Ibai Sistemas, S.A.*
por encargo de: *Subdirección General de Coordinación Bibliotecaria,
Ministerio de Educación, Cultura y Deporte*



Junio de 2.012

© Ministerio de Educación, Cultura y Deporte, 2012



El presente informe pertenece a la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte y está bajo una licencia de ***Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 España*** (<http://creativecommons.org/licenses/by-nc-sa/3.0/es/>) y por ello esta permitido copiar, distribuir y comunicar públicamente esta obra bajo las condiciones siguientes:

Reconocimiento — Debe reconocer los créditos de la obra de la manera especificada haciendo referencia expresa a la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte y al identificador <http://hdl.handle.net/10421/6301> en <http://travesia.mcu.es>. Dicho reconocimiento no podrá en ningún caso sugerir que el Ministerio de Educación, Cultura y Deporte presta apoyo a dicho tercero o apoya el uso que hace de su obra.

No comercial — No puede utilizar esta obra para fines comerciales.

Compartir bajo la misma licencia — Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a ésta.

Para cualquier comentario o sugerencia diríjase por favor a:

Domingo Arroyo Fernández

Subdirección General de Coordinación Bibliotecaria, Ministerio de Educación,
Cultura y Deporte

domingo.arroyo@mcu.es

Javier Garrido

Ibai Sistemas, S.A.

javier@ibai.com

ÍNDICE

1. RESUMEN EJECUTIVO	4
2. REQUISITOS DEL EUROPEANA DATA MODEL	5
3. DESCRIPCIÓN DE LA SOLUCIÓN	6
4. ESQUEMA DE LA SOLUCIÓN	14
5. ENTORNO TECNOLÓGICO	20
6. EQUIPO DE TRABAJO.....	21
7. PLANIFICACIÓN	24
<i>i. Fase de lanzamiento del proyecto.....</i>	<i>25</i>
<i>ii. Fase de análisis de requisitos</i>	<i>25</i>
<i>iii. Fase de diseño de la solución.....</i>	<i>26</i>
<i>iv. Fase de prototipo.....</i>	<i>26</i>
<i>v. Fase de desarrollo e interfaces</i>	<i>27</i>
<i>vi. Fase de implantación en pruebas</i>	<i>27</i>
<i>vii. Fase de validación de la solución.....</i>	<i>28</i>
<i>viii. Fase de carga de datos y validación definitiva</i>	<i>28</i>
<i>ix. Fase de implantación en producción</i>	<i>29</i>
<i>x. Fase de formación</i>	<i>29</i>
<i>xi. Fase de garantía</i>	<i>30</i>
8. TAREAS Y CRONOGRAMA.....	30

1. Resumen ejecutivo

La Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte encargó a la empresa Ibai Sistemas S.A. un estudio sobre cómo realizar la adaptación de DSpace a los requisitos de Europeana. DSpace, una aplicación de gestión de objetos digitales de fuentes abiertas, surgida en el ámbito norteamericano se ha extendido ampliamente a todo el mundo y en España, tal como se reflejaba la encuesta en el presente estudio, cuenta con una comunidad de usuarios y empresas de cierta relevancia; una comunidad que demanda que DSpace converja con el modelo EDM (European Data Model) permitiendo a sus repositorios DSpace interoperar adecuadamente con Europeana.

En la primera parte del estudio realizado por IBAI Sistemas (*Informe de adaptación de DSpace a Europeana – Parte I. Panorama general*¹) se esbozaba una panorámica general de los distintos aspectos del problema.

En esta segunda parte se desarrolla el análisis de una solución técnica detallada de la integración de DSpace y EDM, que servirá de base para los desarrollos de código y elaboración de procedimientos necesarios para la implantación en proyectos reales.

¹ Informe de adaptación de DSpace a Europeana. Fase Danubio: Europeana Data Model (EDM). Disponible en Travesía: <<http://hdl.handle.net/10421/6301>>.

2. Requisitos del Europeana Data Model

Europeana define EDM mediante la siguiente documentación disponible en su web:

- Europeana Data Model Definition v5.2.3²
- Europeana Data Model Primer³
- Europeana Data Model Mapping Guidelines⁴
- Europeana Data Model Factsheet⁵
- Europeana Data Model Presentation (for v5.2.3)⁶

En el “Europeana Data Model Mapping Guidelines” se establece el alcance de la primera implementación de EDM definiendo las siguientes “core classes” (clases principales):

- edm:ProvidedCHO → el objeto de patrimonio cultural
- edm:WebResource → la representación digital del objeto
- ore:Aggregation → la agrupación de las clases

También se definen las “contextual classes” (clases contextuales):

- edm:Agent → quién
- edm:Place → dónde
- edm:TimeSpan → cuándo
- skos:Concept → qué

De forma resumida, estas siete clases son las que DSpace debe cumplir para lograr la adaptación a EDM en su primera implementación.

No todas las propiedades listadas en el “Europeana Data Model Mapping Guidelines” han de ser implementadas inicialmente. Aquellas que no son prioritarias para la primera fase de implantación, están sombreadas en gris en dicho documento.

² EUROPEANA, Europeana Data Model Definition v5.2.3. Disponible en Travesía:
<<http://hdl.handle.net/10421/3861>>.

³ EUROPEANA, Europeana Data Model Primer. Disponible en Travesía:
<<http://hdl.handle.net/10421/5981>>.

⁴ EUROPEANA, Europeana Data Model Mapping Guidelines. Disponible en Travesía:
<<http://hdl.handle.net/10421/5982>>.

⁵ EUROPEANA, Europeana Data Model Factsheet. Disponible en web:
<<http://pro.europeana.eu/documents/900548/f495317b-4557-4a60-9326-723f4618b44c>>.

⁶ EUROPEANA, Europeana Data Model Presentation. Disponible en web:
<<http://pro.europeana.eu/documents/900548/40262c74-d4a3-44d4-95ea-2f4146b50c92>>.

3. Descripción de la solución

Este documento define la solución técnica a aplicar para convertir una instalación DSpace previamente adaptada al esquema Europeana Semantic Elements, al esquema Europeana Data Model según su primera implementación, sentando a su vez las bases para siguientes versiones.

Europeana define que:

- ESE es un subconjunto de EDM.
- ESE puede ser mapeado directamente a EDM.

Así mismo, detalla que esta opción es la que seguirán los 20 millones de registros de que dispone actualmente Europeana.

Para aplicar esta opción propone un mapeo por defecto descrito en la siguiente tabla:

ESE element	Property of provider Aggregation	Property of provider ProvidedCHO
ese:dataProvider	edm:dataProvider	
ese:isShownAt	edm:isShownAt	
ese:isShownBy	edm:isShownBy	
ese:object	edm:object	
ese:provider	edm:provider	
dc:rights	dc:rights	
ese:rights	edm:rights	
ese:type		edm:type
ese:unstored	edm:unstored	
All these remaining Dublin Core properties are mapped to the corresponding Dublin Core property in the edm:ProvidedCHO for that provider. They are therefore properties of the source cultural heritage object. dc:contributor, dc:coverage, dc:creator, dc:date, dc:description, dc:format, dc:identifier, dc:language, dc:publisher, dc:relation, dc:source, dc:subject, dc:title, dc:type, dcterms:alternative, dcterms:spatial, dcterms:extent, dcterms:temporal, dcterms:medium, dcterms:created, dcterms:provenance, dcterms:issued, dcterms:conformsTo, dcterms:hasFormat, dcterms:isFormatOf, dcterms:hasVersion, dcterms:isVersionOf, dcterms:hasPart, dcterms:isPartOf, dcterms:isReferencedBy, dcterms:references, dcterms:isReplacedBy, dcterms:replaces dcterms:isRequiredBy, dcterms:requires dcterms:tableOfContents.		

Las instituciones que disponen de DSpace y deseen cumplir el EDM deberán comprobar los resultados obtenidos en Europeana después de efectuado el mapeo

propuesto y deberán proporcionar un mejor mapeo en caso de que lo consideren oportuno. Lo podrán hacer de varias maneras:

1. Aceptar el mapeo por defecto.

El mapeo por defecto será aplicado por Europeana y las instituciones en este caso no están obligadas a realizar ninguna tarea añadida.

No son necesarias modificaciones en DSpace para los registros existentes en Europeana en caso de elegir esta opción.

Para los registros nuevos, se propone implementar EDM según lo explicado más adelante en el punto 4 “Desarrollar un mapeo completo de datos a EDM”.

2. Crear un nuevo mapeo ESE v3.4.

Con el mapeo por defecto y el resultado de ese mapeo en el portal de Europeana presentes, una institución puede definir un nuevo mapeo ESE v3.4 para proporcionar mejores resultados.

DSpace acepta el almacenamiento de todos los datos presentes en ESE v3.4, así como su recolección por parte de Europeana mediante el plugin ESE, por lo que no son necesarias modificaciones a DSpace.

La tarea de incluir datos diferentes a los ya cargados en DSpace para las etiquetas ESE corresponde a cada una de las instalaciones DSpace, las cuales utilizan para ello programas y reglas de carga propios.

Estos programas de carga son adaptaciones al proceso de importación de DSpace o desarrollos nuevos e independientes realizados por las propias instituciones, los cuales se gestionan de manera autónoma. En cualquiera de los casos disponen de la capacidad de modificarlos según lo que estimen oportuno (según sus orígenes de datos, mapeos desde otros formatos, etc.).

Para los registros nuevos, se propone implementar EDM según lo explicado más adelante en el punto 4 “Desarrollar un mapeo completo de datos a EDM”.

3. Crear un mapeo mínimo a EDM.

Tomando los datos originales que se utilizaron para realizar el mapeo ESE, se puede mapear directamente a EDM, eligiendo exactamente qué dato será aplicado a cada clase.

Para este punto, Europeana proveerá de herramientas de mapeo. Se requerirá la información en formato XML, y habrá que aplicar el mapeo mediante un fichero XSLT.

La estructura del fichero XML será libre y podrá ser definida por cada institución. El fichero XSLT relaciona la información del XML recibida con la de destino (EDM).

Europeana detallará soluciones y ejemplos para los formatos XML más extendidos para bibliotecas, museos y archivos: METS, LIDO y EAD respectivamente.

DSpace posee la capacidad de exportación a METS, por lo que se dispone de la estructura XML en ese formato. Tan sólo falta el fichero XSLT, el cual deberá ser construido. En estos momentos se puede realizar ese fichero XSLT debido a que disponemos de la información de origen (METS en XML) y el formato de destino (EDM, con sus clases y propiedades identificadas y detalladas).

Actualmente existe un único ejemplo sobre cómo un fichero XML LIDO es transformado a EDM. Se prevé que Europeana provea de otros ejemplos para el formato METS. Será entonces el momento más adecuado para acometer esta tarea.

4. Desarrollar un mapeo completo de datos a EDM.

Cuando las instituciones dispongan de datos adicionales que puedan enriquecer el mapeo a EDM, Europeana recomienda utilizar la opción de desarrollar un mapeo completo de datos a EDM.

Esta opción es adecuada, por ejemplo, utilizando las clases contextuales (quién, dónde, cuándo, qué) en las que otras entidades pueden ser modeladas mediante identificadores de recursos contextuales.

DSpace posee limitaciones con respecto a este punto, tanto para los metadatos que exige actualmente la primera implementación de EDM como para los futuros (sombreados en gris en la documentación de Europeana).

Estas limitaciones están relacionadas con dos puntos básicos:

- La forma en que DSpace almacena internamente los datos.
- Las funcionalidades que DSpace proporciona para esos datos.

Para comprender ambos, observamos que las opciones de recuperación de, por ejemplo, los autores (listados, búsquedas, etc.), nos permiten visualizar un autor, pero no acceder a su ficha de metadatos como sí se permite en el caso de un ítem. Esto es debido a que DSpace acepta varios esquemas de

metadatos, pero sólo a nivel de ítem, no de otros elementos como autores y materias.

EDM se basa en RDF, de manera que los recursos deben ser identificados mediante URIs. Esta es la base que permite lograr el Linked Data. En nuestro ejemplo, si nuestros autores en DSpace no disponen de una URI concreta, sólo nos queda el camino de introducir en DSpace una URI externa y en ella (en otra aplicación web diferente) definir los metadatos correspondientes. Aún así, si eligiéramos esta última opción, no tendríamos en DSpace de esos metadatos por lo que no podríamos proporcionárselos a EDM.

La solución a esta problemática se ha encontrado aplicando un plugin a DSpace llamado ASKOSI, el cual sigue la línea de fuentes abiertas y es distribuido mediante licencia GPL.

El plugin ASKOSI dota a DSpace de:

- Una nueva forma de almacenar los datos en Dspace.

A partir de la inclusión de ASKOSI y, siguiendo con nuestro ejemplo, los autores no se almacenarán únicamente en una etiqueta dc.creator o dc.contributor dentro de un ítem, sino que serán ítems en sí mismos. Es decir, los autores serán almacenados como un ítem más, con lo que podremos aprovecharnos de todas las ventajas que ello supone:

- Tendrán una URI persistente, el *handle*, al igual que los ítems, que será gestionado interna y desatendidamente por DSpace.
- Dispondremos de la capacidad para elegir los esquemas de metadatos asociados a los autores.
- DSpace podrá visualizar la descripción de los metadatos asociados.
- Dispondremos de los datos para adaptarlos al formato EDM.

Esta idea se puede extrapolar a otros elementos DSpace que tienen la misma problemática, como por ejemplo las materias.

A continuación mostramos un ejemplo sobre cómo estos datos son almacenados en comunidades/colecciones DSpace para ser gestionados como ítems:

■ 6. Around Musical Works

- 👤 6.1. Authors (Biographies) [19994]
- 📄 6.2. Publishers [1981]
- 🎵 6.3. Orchestra [1735]
- 📖 6.4. Keywords (index) [1019]
- 📁 6.5. Series [1654]

Cada uno de los puntos 6.1 a 6.5 son colecciones DSpace. El punto 6 es una comunidad.

- Funcionalidades específicas para la gestión y visualización de los metadatos.

ASKOSI proporciona una librería jar con una API orientada a los vocabularios controlados. Para ello se basa en RDF y SKOS y proporciona a DSpace la base fundamental sobre la que conseguir almacenar los datos necesarios para lograr EDM, para poder gestionarlos y para hacerlos disponibles mediante OAI-PMH.

Las características principales de ASKOSI API son:

- API alineada con W3C SKOS.
- Orígenes de datos en SQL, CSV, RDF y XML.
- Proporciona estadísticas de uso: qué aplicaciones están utilizando qué conceptos SKOS, cómo (roles) y cuántas veces.
- Está diseñado para el intercambio de datos con otras aplicaciones.

La característica de orígenes de datos en SQL facilita la gestión de los vocabularios controlados, añadiendo una funcionalidad necesaria. Esta característica está desarrollada en DSpace únicamente mediante ficheros estáticos.

ASKOSI dispone a su vez de una parte web que nos permite gestionar este nuevo concepto de almacenamiento de datos en DSpace.

Las características principales de ASKOSI Web son:

- Navegación por listas de autoridades mediante:
 - Árboles de tesauros.
 - Alfabéticamente.

- Por frecuencia de uso.
- SKOS Concepts visualizados en varios formatos.
- Generación de SKOS RDF.
- Validaciones:
 - Errores en los datos.
 - Validaciones de terminología (ambigüedad, traducciones sin realizar).
 - Enlaces a aplicaciones que utilizan SKOS Concepts.

A continuación mostramos un ejemplo sobre cómo ASKOSI gestiona un tesaurus:


The screenshot shows the ASKOSI interface. At the top, there's a 'Scheme list' with options like 'alphabetical or [hierarchy]'. Below it, a search bar contains '1024 concepts'. A sidebar on the left lists various languages: 'Trefwoord', 'Mots-clés', 'Stichwörter', and 'Palabras claves'. The main area displays '1024 concepts' and a table of results. The table has columns for 'Keyword', 'Keyword+NTT', and 'Contains'. The results are listed in a table with 5 columns: a number, a keyword, and three counts.


	Keyword	Keyword+NTT	Contains	
1	Artistic profession	26	104	31
2	ENFANT ARTISTE			1
2	INTERMITTENT DU SPECTACLE			1
2	METIER DE LA DANSE	7	7	3
3	CHOREGRAPHE	5	5	1
3	DANSEUR	2	2	1
3	THEORICIEN DE LA DANSE			1
2	METIER DE LA SCENE	4	1	4


Keywords, A Thesaurus Scheme


Este es un ejemplo sobre cómo se visualiza un autor en DSpace, con su handle 68502/337 en el repositorio del CDMC (Consejo del Condado para la Música y la Cultura de Alta Alsacia)⁷:


⁷ CDMC, (Consejo del Condado para la Música y la Cultura de Alta Alsacia). Jackson, Michael. Disponible en web: <http://www.windmusic.org/dspace/handle/68502/337>


JACKSON, Michael Smk


39

Abstract:  Michael Joseph Jackson (August 29, 1958 – June 25, 2009), known as the King of Pop, was all time. His unique contributions to music and dance, along with a highly publicized personal life in 1971, having made his debut in 1964 with his brothers as a member of The Jackson 5. His 1982 (1987), Dangerous (1991), and HIStory (1995)—among the best selling. He popularized several transformed the music video from a promotional tool into an art form, with videos for his songs crossover following on MTV. His work broke down cultural, racial and generation barriers and Fame, his other achievements feature multiple Guinness World Records—including the Most Successful member of the Jackson 5), and estimated sales between 350 million and 750 million records worth millions of dollars through support of 39 charities and his own Heal the World Foundation. Jack the late 1970s onwards, with changes to his nose and to the color of his skin drawing media pul 2005 he was tried and acquitted of second allegations. He married twice, first in 1994 and again This Is It concert tour in 2009, Jackson died at the age of 50 after suffering from cardiac arrest. ruled a homicide by the Los Angeles County coroner. His memorial service was broadcast live a

 Michael Joseph Jackson (29. August 1958 in Gary, Indiana; 25. Juni 2009 in Los Angeles, Kalifornien) war ein amerikanischer Musiker, Tänzer und Schöpfer von Musikvideos. Er war der kommerziell erfolgreichste Musiker der Welt und hat über 750 Millionen Tonträger verkauft. Jackson war ein Mitglied der Band The Jackson Five. Als Solokünstler hatte er mit den Alben Off the Wall (1979), Bad (1987), Dangerous (1991) und

 Michael Joseph Jackson (Gary, 29 août 1958 – Los Angeles, 25 juin 2009), est un chanteur, danseur et compositeur américain. Il commence sa carrière professionnelle à l'âge de onze ans au sein du groupe The Jackson 5. En 1971, il entame une carrière solo et enregistre 10 albums studio, dont six figurent parmi les plus vendus (1995) et Invincible (2001). Dans les années 1980, Michael Jackson devient une figure majeure de la musique populaire, notamment en concevant des clips musicaux comparables à des courts-métrages de cinéma. Il popularise largement le style de danse du Moonwalk, qui devient sa signature. Surnommé « The King of Pop » (« Le Roi de la pop »), Michael Jackson a battu des records de ventes de disques, ce qui le classe parmi les plus gros vendeurs de disques de tous les temps. Il remporte plus de prix que n'importe quel autre artiste, dont plus de 400 millions de dollars pour des œuvres caritatives, notamment grâce à sa tournée Dangerous. Ses aspects de sa vie privée, notamment son goût pour la chirurgie esthétique, son mode de vie excentrique et ses deux mariages et ses trois enfants furent également à l'origine de polémiques, notamment le jour de sa mort, le 25 juin 2009 au 29 août 2009 les ventes de ses disques sont estimées à plus de 750 millions.

Timeline: 1958 - 2009
Citizenship: United States
Internet Web Site: www.michaeljackson.com/
URI (PermaLink): <http://dspace/handle/68502/337>
Appears in Collections: 6.1. Authors (Biographies)

ASKOSI también permite mejoras notables en DSpace, tanto de cara al usuario como para Europeana.

Una mejora significativa está relacionada con la estructura jerárquica de las materias, lo que permite cierta versatilidad:

Subject Search

Check the boxes next to the categories that you wish to search under, then hit "Search...". Categories can be expanded or collapsed by clicking on the plus or minus sign.

Filtering the list of categories will remove from the list below any categories that do not match the filter term. Expanded categories will be shown in blue.

Find a subject in the controlled vocabulary:

Filter:

- ☒ ☐ Artistic profession (111)
 - ☒ ☐ Casual worker
 - ☒ ☐ Child Artist
 - ☒ ☐ Circus profession (2)
 - ☒ ☐ Dance profession (11)
 - ☒ ☐ Musical profession (71)
 - ☒ ☐ Editing music (10)
 - ☒ ☐ Instrumentalist (11)
 - ☒ ☐ Phonographic production (9)
 - ☒ ☒ Singer (18)
 - ☒ ☐ Stage profession
 - ☒ ☐ Theatre profession (7)
- ☒ ☐ Arts (626)
 - ☒ ☐ Aesthetics (19)
 - ☒ ☐ Architecture (5)
 - ☒ ☐ Fine arts (56)
 - ☒ ☐ Literature (555)
 - ☒ ☐ Photography

ASKOSI es una pieza fundamental para la adaptación de DSpace a EDM.

Queda fuera del alcance de este informe la definición exhaustiva de las modificaciones necesarias a realizar en ASKOSI, debido a que esta tarea forma parte de la ejecución de la solución de DSpace para EDM, y no de su definición. Sin embargo, se muestran las líneas generales de dichas modificaciones:

- Adaptar ASKOSI a la versión de DSpace objeto de adaptación a EDM.
- Adaptar DSpace y ASKOSI para las 7 clases definidas en la primera implementación de Europeana.
- Adaptar el interfaz de DSpace y ASKOSI para la gestión de metadatos de EDM.

Una vez conseguida la estructura de metadatos necesaria para implementar EDM en DSpace junto con las herramientas para gestionarlos y utilizarlos en el repositorio, queda pendiente poner a disposición de los recolectores esta información; para ello se utilizará el protocolo OAI-PMH.

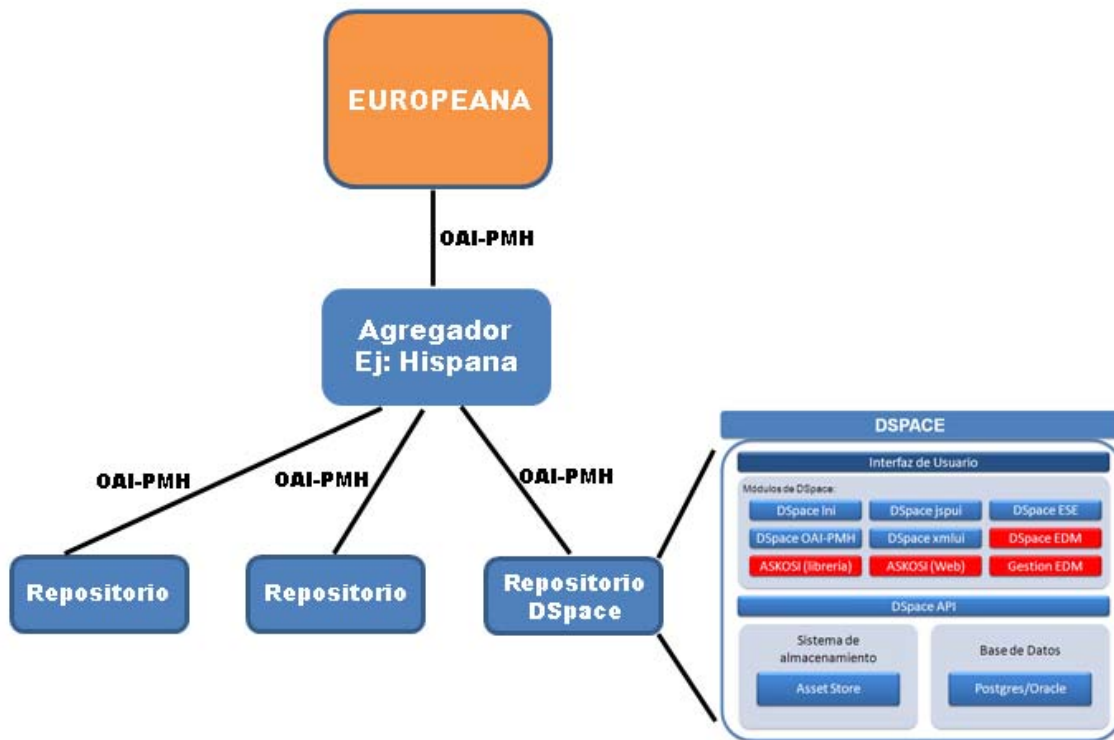
Esta parte se realizará de manera equivalente a como se hizo para el "DSpace plugin for Europeana Semantic Elements", realizado por Vangelis Banos.

Se propone tomar como base dicho plugin ESE y adaptar su programación a las siete clases exigidas por EDM y sus propiedades.

La definición exhaustiva de los cambios a realizar en el plugin es objeto de la ejecución de la solución propuesta en el presente documento, y queda por lo tanto fuera del alcance de este informe.

4. Esquema de la solución

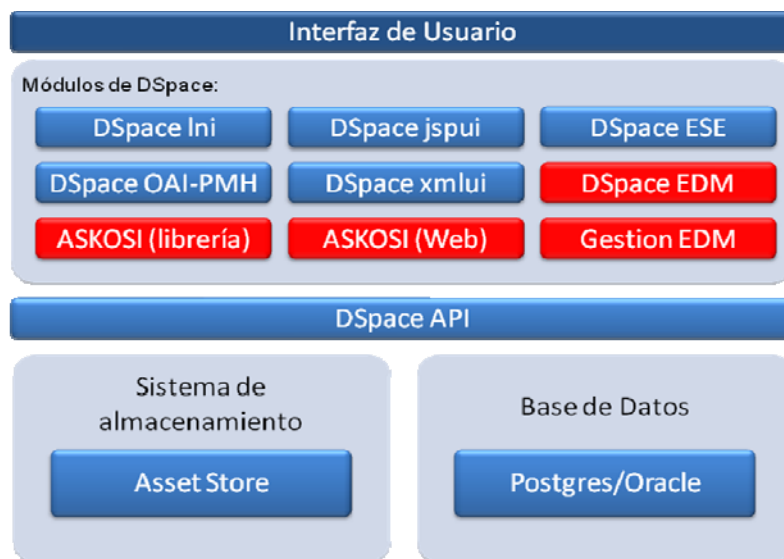
A nivel técnico, la solución a aplicar para adaptar una instalación DSpace al esquema Europeana Data Model se muestra en el siguiente gráfico:



Los repositorios se comunican mediante el protocolo OAI-PMH con los agregadores y estos, a su vez, con Europeana.

En el caso de DSpace y en la actualidad esta operación se realiza mediante el módulo "DSpace ESE".

DSpace dispone de mecanismos para extender sus funcionalidades por medio de nuevos módulos. En el siguiente esquema se observa qué partes son las que hay que desarrollar según la arquitectura DSpace (mostradas en color rojo):



Con los nuevos requisitos del Europeana Data Model se hace necesario:

- Una nueva forma de almacenar los datos en DSpace → mediante ASKOSI (librería) y ASKOSI (Web).
- Dotar a DSpace de funcionalidades requeridas por EDM en su primera implementación → mediante ASKOSI (Web) y el módulo Gestión EDM, que completará los requisitos de Europeana que no cumpla ASKOSI (Web).
- Construir un nuevo módulo tomando como base “DSpace ESE” y extendiéndolo, para proporcionar OAI-PMH → DSpace EDM

El software base necesario que forma parte de la solución técnica se detalla a continuación.

DSpace

La versión actual de DSpace es la versión 1.8.2, sin embargo, para realizar la adaptación a EDM se utilizará la versión 1.7.2. Esto es debido a que la mayoría de las instalaciones Dspace que hay en España corresponden a la versión 1.6.2 o posterior⁸ y porque la versión 1.7.2 es la más estable a día de hoy.

Los requisitos de DSpace 1.7.2 son:

⁸ Informe de adaptación de DSpace a Europeana. Fase Danubio: Europeana Data Model (EDM) – Anexo III, Encuesta de Estado de las instalaciones DSpace. Gráfico “Versión de DSpace” [diciembre de 2011]. Disponible en Travesía: <<http://hdl.handle.net/10421/6301>>.

-
- Sistema operativo UNIX-like OS (Linux, HP/UX, Mac OSX, etc.) o Microsoft Windows.
 - Oracle Java 6 SDK o posterior (SDK es suficiente, no es necesario J2EE).
 - Apache Maven 2.2.x.
 - Apache 1.7 o posterior.
 - Base de datos PostgreSQL y Oracle → se deberá implementar la solución DSpace EDM para ambas.
 - Motor de servlets (Apache Tomcat 5.5 ó 6, Jetty, Caucho Resin o equivalente).
 - Perl (sólo requerido para [dspace]/bin/dspace-info.pl).

ASKOSI

Es un proyecto que consta de herramientas Java que permiten gestionar e integrar vocabularios controlados en otras aplicaciones. Consta de dos partes bien diferenciadas: una librería Java y una aplicación web.

La aplicación web de ASKOSI permite, entre otras funcionalidades, navegar por la lista de autoridades (tesauros), mostrar conceptos SKOS en diferentes formatos, validaciones, etc.

Los requisitos de instalación de ASKOSI son similares a los de DSpace (sistemas operativos y máquina virtual de Java).

Los requisitos de integración de ASKOSI con DSpace son:

- DSpace 1.7.2.
- Base de datos PostgreSQL.

Es necesario, por lo tanto, realizar la migración de la aplicación a la base de datos Oracle.

TAREAS

Los pasos a seguir para la implementación del modelo EDM 5.2.3 son:

Paso 1 - Instalar DSpace

Es necesario disponer de la versión 1.7.2 de DSpace instalada. La base de datos a utilizar será PostgreSQL, debido a que el software desarrollado y resultante de este proyecto será de fuentes abiertas. Posteriormente se realizará la misma operación con la base de datos Oracle.

Paso 2 - Instalar ASKOSI (librería y web)

ASKOSI permite gestionar grupos de datos con formato SKOS. Cada grupo tiene los ficheros de configuración en un directorio. Los ficheros a los que se hace referencia a lo largo de la instalación de ASKOSI se encuentran en:

<http://www.askosi.org/example/>

Paso 3 - Incluir esquemas de metadatos en DSpace

Por defecto DSpace sólo viene configurado con el esquema de metadatos Dublin Core básico <http://purl.org/dc/elements/1.1/>

En el caso del modelo EDM, son necesarios varios esquemas de metadatos más, por lo que hay que incorporarlos a DSpace.

En la siguiente tabla se muestra un listado de los namespaces relacionados con EDM. En cada caso habrá que configurar en DSpace los que van a ser utilizados para cumplir con el modelo EDM.

Nombre	Namespace
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns
owl	http://www.w3.org/2002/07/owl#
dc	http://purl.org/dc/elements/1.1
dcterms	http://purl.org/dc/terms/
ore	http://www.openarchives.org/ore/terms/
foaf	http://xmlns.com/foaf/0.1/
ens	http://www.europeana.eu/schemas/edm/
skos	http://www.w3.org/2008/05/skos
viaf	http://viaf.org/ontology/1.1/
gn	http://www.geonames.org/ontology/

Paso 4 - Construir módulo Gestión EDM para ASKOSI - DSpace

La versión 1.7 y posteriores de DSpace dispone de una API para el control de vocabularios. Esta API es utilizada por ASKOSI para conseguir la integración con DSpace.

Este módulo contendrá las funcionalidades necesarias para gestionar EDM para DSpace y no incluidas en ASKOSI. Estas funcionalidades serán definidas en la fase de análisis detallado del proyecto, en conjunto con la comunidad DSpace y las instituciones implicadas.

Paso 5 - Construir el módulo EDM

Como ocurría para adaptar DSpace al formato ESE, es necesario instalar un plugin para generar los metadatos en formato EDM.

Este plugin que permite generar metadatos mediante el protocolo OAI-PMH en formato EDM debe ser programado específicamente. La base ideal para comenzar dicha programación es utilizar el plugin liberado por Vangelis Banos denominado “DSpace plugin for Europeana Semantic Elements”. Se trata de extender este plugin a partir de DSpace y ASKOSI para lograr EDM.

Paso 6 - Comprobación del cumplimiento de EDM

La comprobación para confirmar que DSpace es capaz de generar formato EDM consiste en visualizar el resultado en la siguiente dirección:

`http://<repositorio>/oai/request?verb=ListRecords&metadataPrefix=edm`

La comprobación real de funcionamiento se realizará validando el resultado obtenido en la dirección anterior con el esquema EDM 5.2.3⁹ publicado por Europeana.

En la actualidad no existe una herramienta de chequeo de EDM, similar a la existente para ESE: Europeana Checker Content Tool¹⁰.

Puntos clave para conseguir la adaptación

La configuración de ASKOSI permite cargar dinámicamente los conceptos de un esquema SKOS, por lo que en el fichero de configuración del esquema que se encuentra en el directorio de datos SKOS, hay que indicárselo con el parámetro `type=DynSQL`.

Lo que nos permite esta carga dinámica es que los conceptos de un esquema se carguen a partir de los elementos que existen en DSpace dentro de una colección específica.

Ejemplo: Para crear un vocabulario controlado de autores, hay que crear en DSpace una colección llamada autores y añadir en esa colección los mismos. Seguidamente es necesario configurar el fichero correspondiente al esquema

⁹ Europeana. EDM XML schema. Disponible en web: `<http://europeanalabs.eu/svn/europeana/trunk/ROOT/src/main/webapp/schemas/edm/>`.

¹⁰ Europeana, *Europeana Content Checker*. Disponible en web: `<http://europeana-contentchecker.isti.cnr.it:8080/launch/>`.

autores en el directorio SKOS para que permita leer la información de dicha colección de DSpace.

Hay que tener en cuenta que a todo elemento creado en DSpace se le asigna un identificador único (*handle*). Teniendo en cuenta esta característica y sabiendo que los metadatos asociados a un elemento de DSpace pueden pertenecer a cualquier namespace (Dublín Core, DCTerms, SKOS, etc...), se consigue una adaptación óptima para el formato EDM.

Resultado final

El presente informe ha definido y descrito la solución técnica a aplicar para la adaptación de DSpace a EDM.

Se propone la realización de un proyecto que construya los módulos informáticos necesarios para conseguir la adaptación a EDM.

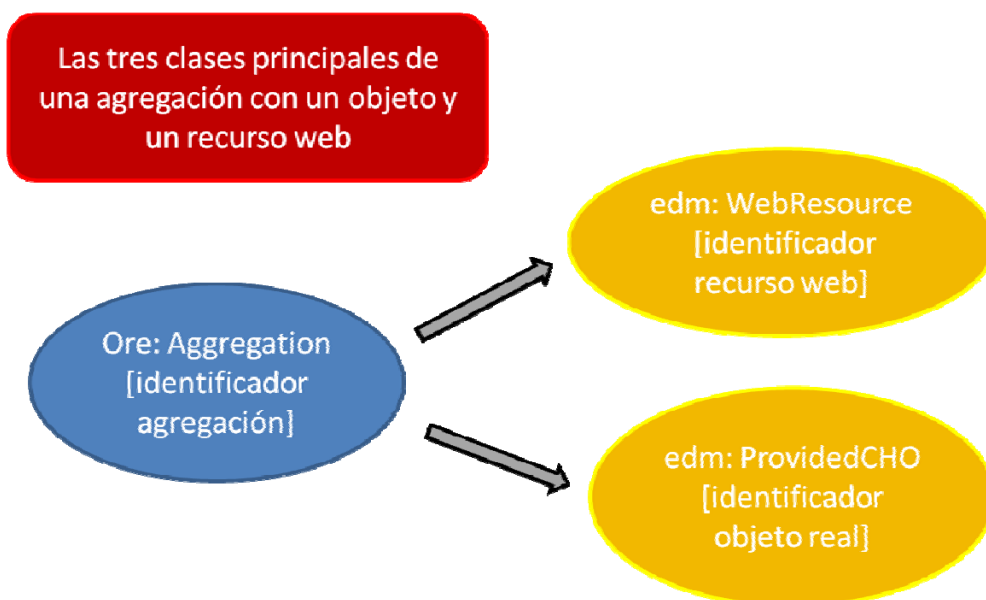
El resultado final de la ejecución de este proyecto propuesto consiste en:

- Generar scripts de migración al nuevo modelo de datos utilizado por ASKOSI.
- Desarrollar los módulos y las modificaciones necesarias: ASKOSI (librería), ASKOSI (web), Gestión EDM y DSpace EDM.
- Documentar la solución.
- Liberar el código fuente.
- Realizar una instalación piloto con DSpace adaptado a EDM.
- Proporcionar soporte a la comunidad DSpace.

Por último, se indica una descripción de las clases principales de EDM:

Cada envío de un ítem a Europeana debería estar creado con 3 tipos de recursos:

1. Objeto origen del patrimonio cultural: se trata del objeto real en sí mismo, es decir, el libro, la imagen, la película, etc. que se está describiendo.
2. Uno de los objetos digitales se presenta: es una o varias representaciones digitales accesibles del objeto, de las cuales algunas pueden ser usadas como pre-visualizaciones de dicho objeto.
3. Un recurso para todo el conjunto: Una agregación que representa el resultado de la actividad del proveedor



Los dos primeros permiten la distinción entre las "obras", que se espera que sea el foco principal de interés de los usuarios, así como sus representaciones digitales, que son los elementos manipulados en los sistemas de información como Europeana.

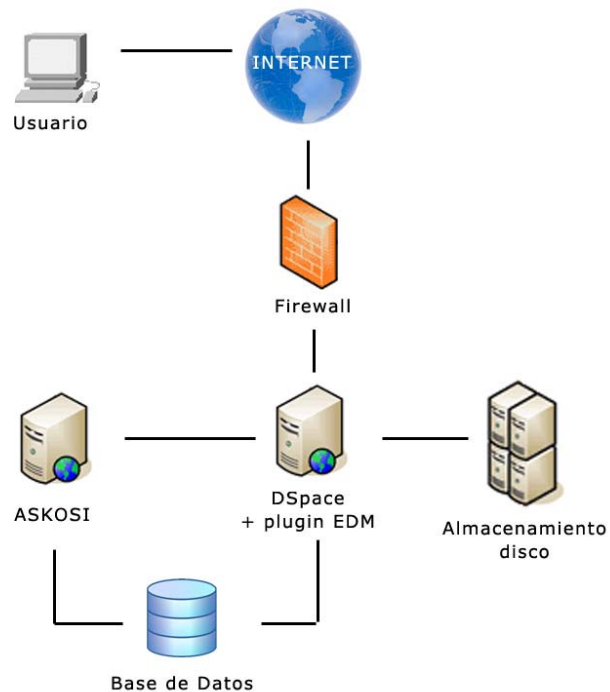
El tercero, siguiendo el enfoque de ORE, demuestra que el objeto proporcionado, junto con las representaciones digitales de un proveedor de datos Europeana puede ser considerada como un conjunto lógico.

En el Informe de adaptación de DSpace a Europeana, Anexo II, Propiedades del Europeana Data Model¹¹, se muestran las propiedades que pueden ser utilizadas para la primera implementación de EDM, tanto de las 3 clases principales como de otras clases. Aparte de las propiedades mostradas, existen más que serán utilizadas en futuras versiones de EDM.

5. Entorno tecnológico

El entorno tecnológico de esta solución se centra en una solución Java, con aplicaciones web (páginas JSP) y bases de datos de fuentes abiertas. El esquema que seguiría esta solución sería similar a la siguiente:

¹¹ Informe de adaptación de DSpace a Europeana. Fase Danubio: Europeana Data Model (EDM) - ANEXO II. Propiedades del Europeana Data Model. Disponible en Travesía: <<http://hdl.handle.net/10421/6301>>.

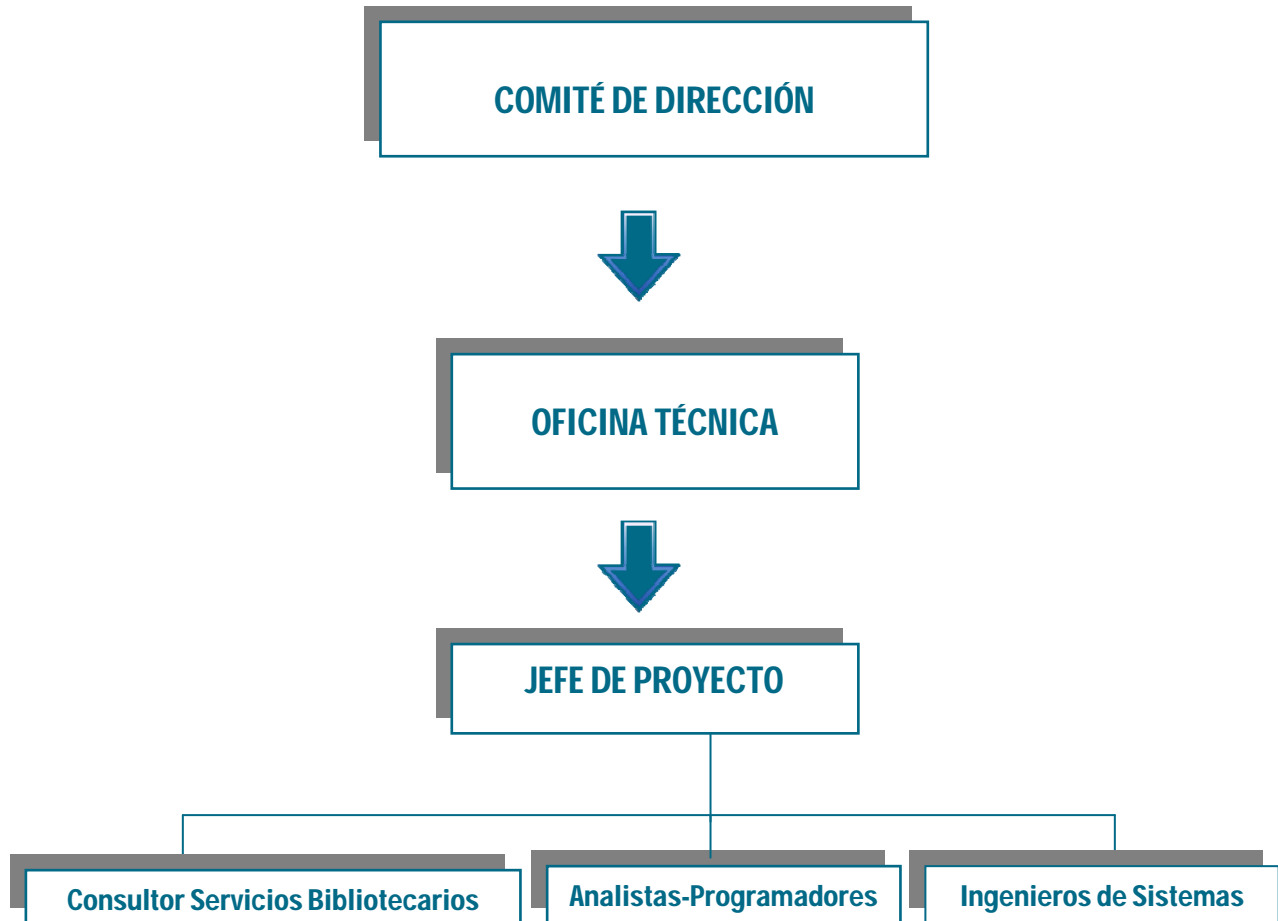


Las aplicaciones web estarían formadas por el software DSpace y ASKOSI que compartirán base de datos (se recomienda el uso de PostgreSQL ya que ambas aplicaciones traen por defecto soporte para esta base de datos).

DSpace tendría acceso a almacenamiento en disco para guardar los archivos visibles desde su interfaz.

6. Equipo de trabajo

A continuación se muestra la forma de organizar el proyecto en cuanto a recursos humanos e infraestructura se refiere. El equipo de trabajo que se considera necesario para el correcto desarrollo del proyecto se representa en el siguiente esquema:



Las funciones del comité de dirección son:

- ⇒ Velar por el cumplimiento de los trabajos exigidos y ofertados.
- ⇒ Emitir las certificaciones parciales de recepción de los mismos.

Las funciones de la oficina técnica de proyecto son:

- ⇒ Control y seguimiento de los trabajos del proyecto, en colaboración con el jefe de proyecto.
- ⇒ Aseguramiento de la disponibilidad de los recursos implicados en el proyecto, según el plan de trabajo.
- ⇒ Aceptación de los productos entregados dentro del proyecto, listos para obtener la certificación del comité de dirección.

Dentro de la oficina técnica las funcionalidades asignadas a cada perfil son:

■ Jefe de proyecto: será el responsable de la correcta ejecución del proyecto, así como de la coordinación de todos los grupos implicados. Las labores asignadas son:

- Definición y evaluación de las necesidades del proyecto.
- Elaborar las especificaciones del proyecto.
- Planificación y control del proyecto.
- Definición del calendario de hitos y entregas.
- Realización de seguimientos e informes de progreso.
- Revisión y evaluación de los resultados.
- Responsable del cumplimiento de los plazos de entrega.
- Dirección del equipo de producción.
- Coordinación del equipo de proyecto.
- Responsable del archivo de la documentación.

■ Consultor en servicios bibliotecarios: es el responsable del asesoramiento en cuestiones específicas relacionadas con los servicios bibliotecarios, con las siguientes tareas:

- Identificación de necesidades bibliotecarias.
- Labores de consultoría bibliotecaria en los puntos que el jefe de proyecto estime oportunos.
- Estructuración de datos acorde a la documentación actual.
- Definición de perfiles de usuarios y control de acceso.
- Elaboración del manual de procedimiento, si procede.
- Estructuración lógica de la información.

■ Analista-programador: este rol está encargado de interpretar y desarrollar las nuevas funciones y modificaciones, tanto desde el punto de vista funcional, como desde el de la evolución y mejoras:

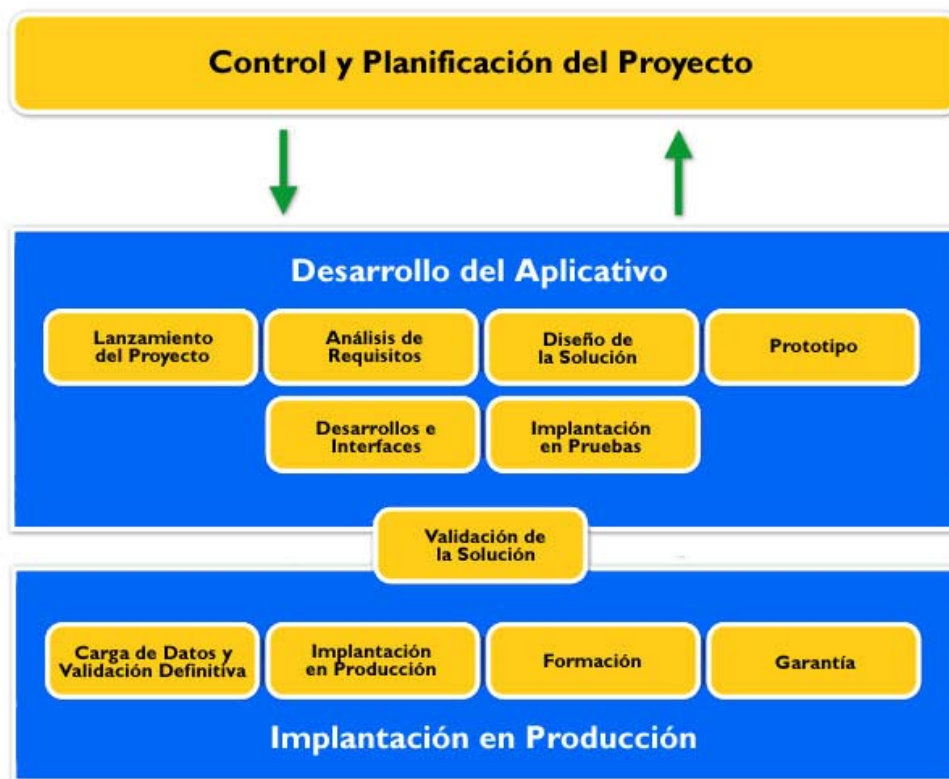
- Elaboración de los requisitos que definen el sistema.
- Elaboración de los modelos de clases e interacción de objetos.
- Especificación de la interfaz del sistema.
- Construcción y mantenimiento del sistema.
- Generación del código asociado a los procedimientos a desarrollar.
- Participación en las pruebas de conjunto.
- Conocimiento del uso de librerías y frameworks.

■ Ingeniero de sistemas: este rol está encargado de instalar y operar las instalaciones de hardware y software:

- Ejecución del plan de implantación de sistemas.
- Definición de la arquitectura del sistema.
- Instalación y configuración del entorno tecnológico.
- Diseño físico de los datos.
- Migración de la estructura y carga inicial de datos.
- Mantenimiento del sistema.
- Responsable de la seguridad de sistema.
- Soporte técnico al resto de integrantes del equipo de trabajo.

7. Planificación

En el siguiente diagrama se representan, de forma esquemática, las diferentes fases que se llevarán a cabo a lo largo del proyecto:



Seguidamente se describe de forma detallada cada una de ellas:

i. Fase de lanzamiento del proyecto

Lanzamiento del Proyecto	
Objetivos	
<ul style="list-style-type: none">✓ Presentar el equipo de implantación✓ Definición de la plataforma de biblioteca digital: DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS✓ Confirmar el alcance, objetivos y prioridades✓ Planificar el proyecto	
Actividad	Resultados
<ul style="list-style-type: none">◆ Presentación del jefe de proyecto◆ Presentación del repositorio OAI DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS al personal implicado en el proyecto◆ Definición de puntos críticos	<ul style="list-style-type: none">◆ Objetivos y alcance del proyecto fijados◆ Planificación de los hitos del proyecto◆ Equipo constituido y formado◆ Primera toma de contacto entre los implicados en el proyecto realizada◆ Puntos críticos en el desarrollo del proyecto identificados por ambas partes

ii. Fase de análisis de requisitos

Análisis de Requisitos	
Objetivos	
<ul style="list-style-type: none">✓ Concretar las necesidades funcionales del sistema✓ Presentar el catálogo de requisitos y relación de módulos del sistema	
Actividad	Resultados
<ul style="list-style-type: none">◆ Análisis de la situación de las instalaciones DSpace en España y de sus necesidades funcionales◆ Análisis de las aplicaciones (ASI)	<ul style="list-style-type: none">◆ Conocimiento del marco de trabajo y funcionalidades básicas◆ Arquitectura del sistema

iii. Fase de diseño de la solución

Diseño de la Solución	
Objetivos	
<ul style="list-style-type: none">✓ Definición detallada de los procesos a utilizar y de los flujos de información necesarios✓ Diseño de las aplicaciones (DSI)✓ Identificación de las adaptaciones necesarias	
Actividad	Resultados
<ul style="list-style-type: none">◆ Definir cómo el nuevo sistema soportará los procesos de trabajo diario◆ Definir los posibles desarrollos a realizar◆ Plantear diferentes alternativas a los puntos críticos del sistema	<ul style="list-style-type: none">◆ Fijación de un calendario de trabajo◆ Requisitos detectados y valorados◆ Actualización de la planificación para el resto de fases posteriores

iv. Fase de prototipo

Prototipo	
Objetivos	
<ul style="list-style-type: none">✓ Contrastar las necesidades del sistema con la aplicación diseñada✓ Elaborar un prototipo que sirva como base de trabajo para las siguientes fases✓ Instalación del prototipo en un servidor de pruebas para próximos trabajos de desarrollo✓ Propuesta gráfica para del repositorio OAI	
Actividad	Resultados
<ul style="list-style-type: none">◆ Instalación del repositorio DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS básicos en un servidor de pruebas◆ Introducción de un volumen de datos significativo para su utilización◆ Estudiar el prototipo y validarlo◆ Definir el entorno gráfico	<ul style="list-style-type: none">◆ Verificación de los puntos críticos detectados en fases anteriores◆ Prototipo validado◆ Prototipo preparado para desarrollo en un entorno de pruebas

v. Fase de desarrollo e interfaces

Desarrollos e interfaces	
Objetivos	
<ul style="list-style-type: none">✓ Realización de los desarrollos necesarios✓ Realización de las integraciones necesarias	
Actividad	Resultados
<ul style="list-style-type: none">◆ Trabajos realizados sobre plataforma de desarrollo	<ul style="list-style-type: none">◆ Fin de los trabajos de desarrollo◆ Interfaces de trabajo construidas

vi. Fase de implantación en pruebas

Implantación en pruebas	
Objetivos	
<ul style="list-style-type: none">✓ Implantación en pruebas de las personalizaciones y desarrollos realizados sobre el repositorio OAI DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS estándar✓ Verificación de la integración del sistema	
Actividad	Resultados
<ul style="list-style-type: none">◆ Sistema implantado en entorno de Pruebas◆ Entrega del manual técnico de DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS	<ul style="list-style-type: none">◆ Verificación del sistema en entorno de pruebas

vii. Fase de validación de la solución

Validación de la solución	
Objetivos	
✓ Obtener la aprobación formal del nuevo repositorio OAI DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS de acuerdo con los requerimientos funcionales, de organización y de integración con los otros sistemas	
Actividad	Resultados
◆ Verificación del funcionamiento con los datos del prototipo	◆ Validación del sistema en entorno de Pruebas

viii. Fase de carga de datos y validación definitiva

Carga de datos y validación definitiva	
Objetivos	
✓ Funcionamiento con datos reales ✓ Recepcionar la solución	
Actividad	Resultados
◆ Carga inicial de datos ◆ Funcionamiento del sistema con el volumen de datos real	◆ Confirmación del modelo funcional obtenido en la fase de diseño de la solución ◆ Sistema validado para el arranque

ix. Fase de implantación en producción

Implantación en producción	
Objetivos	
<ul style="list-style-type: none">✓ Puesta en marcha del repositorio OAI DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS en una instalación en Producción✓ Traspaso a la comunidad DSPACE + ASKOSI + DESARROLLOS ESPECIFICOS de información sobre el repositorio OAI✓ Soporte a usuarios finales✓ Aceptación final del funcionamiento del sistema	
Actividad	Resultados
<ul style="list-style-type: none">◆ Sistema implantado en entorno de producción◆ Sistema integrado con el resto de aplicaciones◆ Entrega de código fuente de la aplicación y scripts de datos	<ul style="list-style-type: none">◆ Sistema operacional con datos reales◆ Sistema en entorno productivo

x. Fase de formación

Formación	
Objetivos	
<ul style="list-style-type: none">✓ Formar a los administradores del sistema para que sean capaces de colaborar en la implantación✓ Formar a los usuarios finales para que sean capaces de utilizar el sistema de manera autónoma✓ La formación se realizará <i>in-situ</i>✓ Las jornadas de formación se complementan con trabajo personal	
Actividad	Resultados
<ul style="list-style-type: none">◆ Entrega del manual de instalación / explotación (soporte papel y electrónico)◆ Entrega de los manuales de usuario (soporte papel y electrónico)◆ Entrega del informe de pruebas unitarias y de integración	<ul style="list-style-type: none">◆ Plan de formación elaborado◆ Administradores formados◆ Usuarios finales formados

xi. Fase de garantía

Para el aseguramiento de la calidad en la ejecución de los trabajos, se exige la realización de las siguientes actividades:

- Control de calidad durante todas las fases y actividades del proyecto.
- Revisión y control de los hitos y entregables.
- Revisión y control de calidad previa a la entrega de documentación final y cierre de los proyectos.

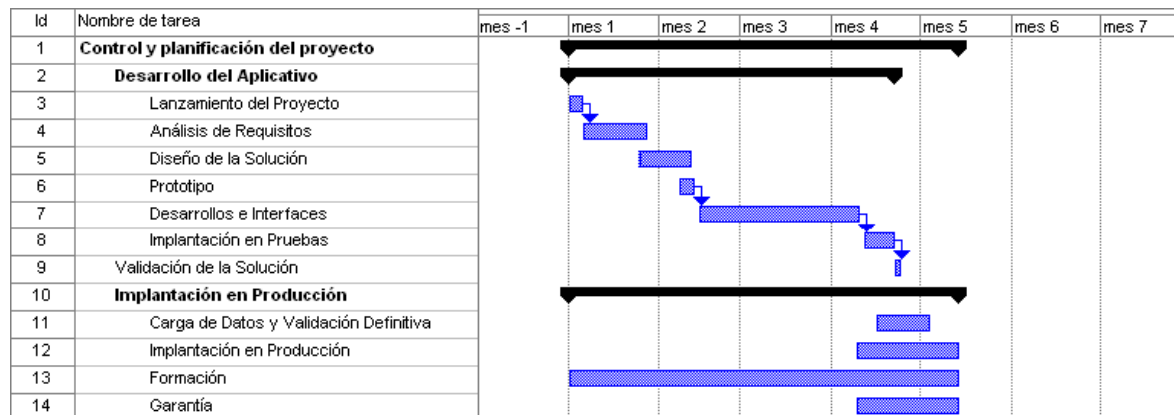
A tal efecto, se establece la necesidad de disponer de un certificado de calidad a nivel de empresa, del tipo ISO o similar, cuya metodología será aplicada a este proyecto.

8. Tareas y cronograma

Las fases/tareas del proyecto son las siguientes:

Id.	FASE
DSP-LPR	Lanzamiento del proyecto
DSP-ASI	Análisis de requisitos
DSP-DSI	Diseño de la solución
DSP-PRO	Prototipo
DSP-CSI	Desarrollos e interfaces
DSP-IASa	Implantación en pruebas
DSP-IASb	Validación de la solución
DSP-IASc	Carga de datos y validación definitiva
DSP-IASd	Implantación en producción
DSP-FOR	Formación
DSP-GAR	Garantía

El siguiente esquema muestra los tiempos de ejecución en cada una de las fases del proyecto:



El desglose de horas por perfil y fase del proyecto es el siguiente:

Descripción Perfil	Gestión Proyecto		Análisis		Diseño		Construcción		Implantación		Total Horas
	Horas	%	Horas	%	Horas	%	Horas	%	Horas	%	
Jefe de proyecto	40	4%	40	4%	40	4%	40	4%	40	4%	200
Analista-programador			60	6%	100	10%	500	48%			660
Consultor bibliotecario			100	10%							100
Ingeniero de sistemas									80	8%	80
TOTAL	40		200		140		540		120		1.040

El presente informe es el resultado de haber completado la fase de lanzamiento del proyecto, y de haber realizado las fases de análisis de requisitos y diseño de la solución en su parte de definición funcional general. A partir de este punto es necesario profundizar en la definición técnica de dichas fases, con el objetivo de implementar el prototipo de la solución.